Journal of Statistical Planning and Inference 153 (2014) 1–10

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

Some asymptotic properties for functional canonical correlation analysis

Heng Lian

School of Mathematics and Statistics, University of New South Wales, Sydney, NSW, 2052, Australia

ARTICLE INFO

Article history:
Received 1 November 2013
Received in revised form
11 April 2014
Accepted 24 May 2014
Available online 5 June 2014

Keywords:
Canonical correlation analysis
Covariance operator
Cross-covariance operator
Lower bound

ABSTRACT

We consider convergence rates of functional canonical correlation analysis (FCCA). There are already several studies on FCCA in the literature, which focused on its population properties as well as consistency. Our setup most closely resembles that of He et al. (2003). Under an assumption that controls the level of dependence (roughly that the dependence between the two functional objects is not too high), we derive convergence rates of the weight functions to their population counterpart. Both upper bound and lower bound are derived for the L^2-norm and the prediction risk (also called Σ-norm) of the weight functions.

1. Introduction

Canonical correlation analysis (CCA), introduced by Harold Hotelling (Hotelling, 1936), seeks linear combinations of $\mathbf{x} \in R^p$ and $\mathbf{y} \in R^q$ that have maximum possible correlation. In the population, CCA solves the problem $\max_{\mathbf{u},\mathbf{v}: Var(\mathbf{u}^T\mathbf{x}) = Var(\mathbf{v}^T\mathbf{y}) = 1} Cov(\mathbf{u}^T\mathbf{x}, \mathbf{v}^T\mathbf{y})$ to get the first pair of weight vectors. Subsequent weight vectors can be extracted sequentially by adding some orthogonality constraints with the previously extracted ones.

There are increasingly more cases in practice where data sets are collected in which the basic units of measurements appear as curves. Functional data analysis, as an important subdiscipline of statistics as of today, is developed to handle this situation. Many statistical tools from multivariate analysis have been extended, including linear and nonlinear regression (Cardot et al., 1999; Yao et al., 2005b; Mueller and Stadtmueller, 2005; Yao and Mueller, 2010, nonparametric and semiparametric regression (Ferraty and Vieu, 2002, 2004; Preda, 2007; Mueller and Yao, 2008; Chen et al., 2011), principal component analysis (Silverman, 1996; Hall and Mohammad, 2006; Yao et al., 2005a), classification (Hall et al., 2001; Ferraty and Vieu, 2003, 2004), and clustering (James and Sugar, 2003; Ray and Mallick, 2006; Chiou and Li, 2007; Ma and Zhong, 2008; Jacques and Preda, 2014). In particular, functional version of CCA has been studied in Leurgans et al. (1993), where the motivating example given is the data on children's gait to study the relationship between knee joint angle and hip joint angle. Another example of using FCCA is in the study of functional connectivity of the brain on MEG or fMRI data. Subsequent developments in FCCA have contributed to many theoretical and computational aspects of the method. For example He et al. (2003) clarified the well-definedness of FCCA in the population and He et al. (2004) proposed several computational methods. Fukumizu et al. (2007) rigorously proved statistical consistency of CCA in the framework of reproducing kernel Hilbert space which is slightly more general than FCCA. Based on the observation that weight functions in FCCA may not always exist (the maximum value is not achieved by any pair even though the value is well-defined),

E-mail address: henglian@ntu.edu.sg

http://dx.doi.org/10.1016/j.jspi.2014.05.008
0378-3758/

Cupidon et al. (2008) suggested using regularized FCCA even in the population to avoid this problem. Eubank and Hsing (2008) further proposed a very general framework that does not even require the functional variables to be in L^2 and studied its consistency. Finally we mention the more recent work of Yang et al. (2011) that redefined the FCCA to avoid the inverse problem, which we prefer to call functional singular component analysis instead of FCCA (as do the authors for that paper). In Yang et al. (2011), since the inverse operator is entirely avoided, theoretically and computationally it becomes simpler (in particular no regularization is required).

The current paper is solely focused on minimax convergence rates of FCCA under the setup of He et al. (2003) which puts the functional variables inside $L^2[0,1]$ (we will assume that the functions are defined on the unit interval [0,1] without loss of generality). However, as far as we know, convergence rate for FCCA has never been established in the literature. It is of significant interest to understand how well the weight functions can be estimated in the minimax sense as a benchmark of estimation performance, and with convergence defined in different distances. Given the complicated form of the estimator which is defined as an eigenfunction of some operator making it different from many existing studies of minimax estimation in nonparametric regression, this is a challenging task. To obtain nontrivial convergence rate, we will impose stronger assumptions on the covariance and cross-covariance operators. On the other hand, under these stronger assumptions, the maximum value in FCCA is always well-defined and achieved by a pair of weight functions.

The rest of the paper is organized as follows. In Section 2, we review the population setup. We also introduce conditions under which and weight functions are well defined, which are mostly based on He et al. (2003). Section 3 contains the main theoretical results on convergence rates of the weight functions, showing the upper bound in both L^2-norm and prediction risk (Σ-norm). Section 4 considers derivations of the lower bounds. Section 5 presents a simulation example. Section 6 concludes the paper with a discussion.

2. Functional canonical correlation analysis

In the FCCA problem, we assume X and Y are square integrable random processes indexed by $t \in [0,1]$ which is denoted simply by $X, Y \in L^2[0,1]$. As always assumed in the functional linear regression literature (Cardot et al., 1999; Yao et al., 2005b; Cai and Hall, 2006; Hall and Horowitz, 2007), we assume $E\|X\|^4, E\|Y\|^4 < \infty$ where $\|X\| = (\int_0^1 X^2)^{1/2}$ is the L^2 norm of X for example. We also use $\langle \cdot \rangle$ to denote the standard inner product in L^2. Note that when studying the population properties, only $E\|X\|^2, E\|Y\|^2 < \infty$ is necessary. Fourth moment is required when studying convergence rates (see for example assumption (C3) in the next section). Without loss of generality, we assume that the variables are centered with $EX = EY = 0$. For use in FCCA, we define the covariance operators $\Sigma_X = E[X \otimes X]$, $\Sigma_Y = E[Y \otimes Y]$, which are trace-class operators under the assumption $E\|X\|^2, E\|Y\|^2 < \infty$. Remember that for $f, g \in L^2[0,1]$, $f \otimes g$ is the linear operator that maps $h \in L^2[0,1]$ to $\langle g, h \rangle f \in L^2[0,1]$. As usual we assume that $Ker(\Sigma_X) = Ker(\Sigma_Y) = \{0\}$. Without this assumption we will simply need to restrict consideration of weight functions to the subspaces $Ker(\Sigma_X)^\perp$ and $Ker(\Sigma_Y)^\perp$. We also define the cross-covariance operators $\Sigma_{XY} = E[X \otimes Y]$ and $\Sigma_{YX} = \Sigma_{XY}^T = E[Y \otimes X]$ where $()^T$ denotes the conjugate operator.

Under the assumption $E\|X\|^2, E\|Y\|^2 < \infty$, by Mercer's theorem, Σ_X, Σ_Y can be expressed as

$$\Sigma_X = \sum_{j=1}^{\infty} \lambda_{Xj} \phi_j \otimes \phi_j, \quad \Sigma_Y = \sum_{j=1}^{\infty} \lambda_{Yj} \psi_j \otimes \psi_j,$$

where $\lambda_{X1} \ge \lambda_{X2} \ge \cdots > 0$ and $\lambda_{Y1} \ge \lambda_{Y2} \ge \cdots > 0$ are the eigenvalues, and $\phi_j, \psi_j \in L^2[0,1], j = 1, 2, \ldots,$ are the orthonormal eigenfunctions. Correspondingly, we have the Karhunen–Loève expansions for the random processes, given by

$$X = \sum_{j=1}^{\infty} \xi_j \phi_j, \quad Y = \sum_{j=1}^{\infty} \eta_j \psi_j.$$

The expansions above only depend on the marginal distributions of X and Y. In FCCA, we pose the maximization problem

$$\rho = \max_{(f,g):\langle f, \Sigma_X f \rangle = \langle g, \Sigma_Y g \rangle = 1} \langle f, \Sigma_{XY} g \rangle. \tag{1}$$

Here we only focus on the first pair of weight functions, with the understanding that subsequent pairs can be extracted sequentially. Also note that we do not distinguish between maximum and supremum in our mathematical expressions. That is, the maximum value above may not be achieved in general. Nevertheless, we will impose stronger assumptions soon that get rid of this problem.

To make the maximum value in (1) achievable, we impose the following sufficient (although not necessary) condition which is the same as Condition 4.5 in He et al. (2003).

(C1) Denote $e_{ij} = E[\xi_i \eta_j]$. We have $\sum_{i,j=1}^{\infty} \frac{e_{ij}^2}{\lambda_{Xi}^2 \lambda_{Yj}} < \infty$ and $\sum_{i,j=1}^{\infty} \frac{e_{ij}^2}{\lambda_{Xi} \lambda_{Yj}^2} < \infty$.

Proposition 1. *Under condition (C 1), the operators $\Sigma_X^{-1/2} \Sigma_{XY} \Sigma_Y^{-1/2}$ and $\Sigma_X^{-1} \Sigma_{XY} \Sigma_Y^{-1/2}$ are Hilbert-Schmidt operators defined on L^2 (by continuity extension). The maximum in (1) is achieved for some pair of weight functions (f,g).*